Original Research Article

Modeling COVID-19 pandemic using Bayesian analysis with application to Slovene data

Damjan Manevski, Nina Ružić Gorenjec, Nataša Kejžar, Rok Blagus *

Institute for Biostatistics and Medical Informatics, Faculty of Medicine, University of Ljubljana, Vrazov trg 2, 1000 Ljubljana, Slovenia

ARTICLE INFO

ABSTRACT

In the paper, we propose a semiparametric framework for modeling the COVID-19 pandemic. The stochastic part of the framework is based on Bayesian inference. The model is informed by the actual COVID-19 data and the current epidemiological findings about the disease.

The framework combines many available data sources (number of positive cases, number of patients in hospitals and in intensive care, etc.) to make outputs as accurate as possible and incorporates the times of non-pharmaceutical governmental interventions which were adopted worldwide to slow-down the pandemic. The model estimates the reproduction number of SARS-CoV-2, the number of infected individuals and the number of patients in different disease progression states in time. It can be used for estimating current infection fatality rate, proportion of individuals not detected and short term forecasting of important indicators for monitoring the state of the healthcare system. With the prediction of the number of patients in hospitals and intensive care units, policy makers could make data driven decisions to potentially avoid overloading the capacities of the healthcare system. The model is applied to Slovene COVID-19 data showing the effectiveness of the adopted interventions for controlling the epidemic by reducing the reproduction number of SARS-CoV-2. It is estimated that the proportion of infected people in Slovenia was among the lowest in Europe (0.350%, 90% CI [0.245–0.573]%), that infection fatality rate in Slovenia until the end of first wave was 1.56% (90% CI [0.94–2.21]%) and the proportion of unidentified cases was 88% (90% CI [83–93]%).

The proposed framework can be extended to more countries/regions, thus allowing for comparison between them. One such modification is exhibited on data for Slovene hospitals.

1. Introduction

The new coronavirus SARS-CoV-2 has quickly spread all around the world greatly impacting all aspects of our lives. One of the key reasons for its rapid spread is the high reproduction number of infection, R_t. The R_t value represents the average number of people that an individual infects during their own infection period, whereby t represents time. R_t is susceptible to change due to interventions, social behavior, etc. When $R_t < 1$, the incidence of new cases reduces, but when $R_t > 1$, it increases until the epidemic reaches its peak; after that, the incidence of new cases starts to reduce due to (at least temporal) herd immunity [1]. The estimates of the basic reproduction number, R_0, for SARS-CoV-2 vary substantially, depending on the method of estimation, and stand at about 3 [2–8]. Such a high basic reproduction number leads to a steep exponential increase in the number of cases which, in turn, causes a rapid increase in the number of people in need of hospitalization and intensive care unit (ICU). Due to the limited capacities of the healthcare system, this can lead to a situation where it is impossible to provide suitable care to all patients in need. Thus, it is of crucial importance for policy makers to estimate R_t, infection fatality rate, proportion of asymptomatic cases and forecasting the number of hospitalized patients and patients in the ICUs.

In the current COVID-19 pandemic many governments have adopted non-pharmaceutical interventions (NPIs) to control the spread of the epidemic in their countries. Many models are available for forecasting the COVID-19 pandemic with adopted NPIs, e.g. SIR-like (Susceptible–Infected–Recovered-like) compartmental models that are essentially deterministic models of stochastic processes, which capture the mean dynamic quite well for large populations [9–14]. Another possible approach is using large simulation models that include networked (meta)populations or are agent-based and thus mimic the behavior of all individuals [11,15,16]. Xue et al. [14] developed a network-based model extending the network SIR model to capture the contact heterogeneity among individuals.

Bayesian inference was introduced as an alternative to the compartmental models [17–19]. Here, simple stochastic models which describe

the key features of epidemic spread are formulated and then actual data are used to estimate the parameters of the model, thus allowing the analysis to be driven by data and using less assumptions and/or fixed parameters [20]. Furthermore, the final estimates are equipped with statistically sound intervals of uncertainty.

Flaxman et al. [7] studied the influence of NPIs relying on Bayesian inference, using data for 11 European countries. They concluded that, after adopting all interventions, the reproduction number of infection decreased from the basic value of 3.8 (mean value for all 11 countries) to 0.66 (ranging from 0.44 in Norway to 0.82 in Belgium). In their work, only country data on daily number of deaths (up to and including May 4, 2020) were taken into consideration. The same model was applied by also including Slovenia in the analysis (using data up to and including April 12, 2020), showing that NPIs were effective also in Slovenia, reducing the reproduction number to 0.6 with the whole 95% credible interval being below 1 [21]. One of the main goals of these analyses was to accurately estimate the cumulative effect of interventions while using only data on number of deaths. For this reason, it was crucial to pool information across different countries whereby imposing some strong assumptions. Namely, it had to be assumed that the reproduction number of infection does not change between given interventions and that their effect is similar across countries. Therefore, the results based on the Flaxman model [7] are driven more by countries with advanced epidemics and larger number of deaths [7]. E.g., excluding the data for Italy and Spain, the estimate of the reproduction number for Slovenia after adopting the NPIs changes to 0.8 with 95% credible interval including 1 [21]. Additionally, there is a large time lag in deaths data which may inform the parameters of the model. Furthermore, the Flaxman model [7] cannot be directly used for forecasting the number of hospitalized patients and patients in ICUs or to estimate the proportion of unidentified cases.

We therefore propose an elaborated semiparametric modeling framework based on a Bayesian approach, that greatly extends and modifies the Flaxman model. In contrast to Flaxman et al. [7] which only uses daily number of deaths, we propose an extension using additional data sources for important indicators of the state of epidemic and the healthcare system that are crucial for effectively coping with the pandemic and have less 'time lag'. We use: confirmed number of cases and deaths, number of hospitalized patients, number of patients in ICUs, number of patients admitted to and released from hospitals and ICUs.

The rationale for including more data for a given country is manyfold: (1) to obtain accurate estimates for countries with small number of deaths and/or which are at the earlier stages of the epidemic without having to pool the information from other countries, (2) to provide estimates of more parameters, which allow to report other measures, such as proportion of unidentified cases, (3) to make it possible to relax some strong assumptions around NPIs (i.e. to address critiques about the piece-wise constant R_t of the Flaxman model).

We model the time varying effective reproduction number differently than the Flaxman model allowing more flexibility under weaker assumptions. The outputs of our model are estimates of R_t and other parameters with narrower credible intervals due to the inclusion of different data sources, and accurate short term forecasts of various important epidemic indicators. The quality of the model and its forecasts is evaluated on real data.

We apply the proposed modeling framework using publicly available data for Slovenia [22]. We estimate the reproduction number of SARS-CoV-2 and provide accurate short term forecasts for important indicators for monitoring the state of the healthcare system throughout the course of the Slovene COVID-19 epidemic. We show how the proposed framework can be used to estimate the effectiveness of the adopted NPIs (and their lifting) for Slovenia and estimate some measures at the final time point, i.e., infection fatality rate, proportion of cumulatively infected people and proportion of unidentified cases. Flexibility of the proposed modeling framework is also illustrated using

data at hospital level, providing separate forecasts for each hospital where COVID-19 patients were admitted.

The rest of the article is organized as follows. In the next section we formally present the model which is applied to Slovene COVID-19 data and show its possible extensions/generalizations. This is followed by showing a series of selected results for Slovene epidemic. The paper concludes with a discussion and a summary of the most important findings.

2. Modeling framework

Each country has a slightly different set of data, which usually does not fit to a model that is optimal from the epidemiological point of view. The main idea is to develop a model that is applicable to available data, adapting it to suit Slovene data, but making it easy to adjust according to availability and reliability of data in any country. First, we present the core model for Slovenia. Then, we show how it may be extended or modified to a more general framework. Finally, we illustrate one possible modification of the model.

We adopt the following notation. We add its meaning in parentheses so that the notation is easier to follow. Daily data on the number of confirmed cases is denoted by P_t (positive cases), number of hospitalized patients by H_t, number of patients in Intensive Care Units (ICU) by U_t, daily number of new patients admitted to hospitals and to ICUs by H_t^I (hospital in) and U_t^I (ICU in), respectively, daily number of patients released from hospitals and ICUs by H_t^O (hospital out) and U_t^O (ICU out), respectively, daily number of deaths occurring in hospitals by D_t^H (death from hospital), and daily number of deaths occurring outside the hospitals by D_t^C (death directly from cases), where $t = 1, \ldots, n$ are days from the start until the end of the analysis.

Let $\mathcal{N}(\mu, \sigma)$ denote a normal distribution with mean μ and standard deviation σ. We say that X follows a folded-normal distribution, denoted by $\mathcal{N}^+(\mu, \sigma)$, if $X = |Y|$ and $Y \sim \mathcal{N}(\mu, \sigma)$. Let $\text{Unif}(a, b)$ denote a continuous uniform distribution on $[a, b]$, let $\Gamma(\mu, \xi)$ denote a gamma distribution with mean μ and coefficient of variation ξ, and let $\text{Exp}(\lambda)$ be the exponential distribution with mean $1/\lambda$. Using $\text{NegBin}(\mu, \mu + \frac{\mu^2}{\phi})$ we denote the negative binomial distribution with mean μ and variance $\mu + \frac{\mu^2}{\phi}$. If f is the density of a positive random variable, then $\{f_s\}_{s=1}^{\infty}$ denotes its discretization via

$$f_1 = \int_0^{1.5} f(t)dt, \quad f_s = \int_{s-0.5}^{s+0.5} f(t)dt \quad \text{for } s = 2, 3, \ldots \quad (1)$$

2.1. The model

The core model is schematically presented in Fig. 1. We explain it thoroughly in the next subsections. First, we present the upper part (see line border on Fig. 1) that is used for modeling infections, then the lower part used for modeling disease progression, and in the last subsection the whole model is precisely specified. In Fig. 1, the data are presented in circles, the deterministic parameters of the model are colored red, while 'true' parameters of the model, for which posterior distributions are estimated, are in blue (the parameters with prior distributions) or black (the transformed parameters).

2.1.1. Modeling infections

Following Flaxman et al. [7], the number of truly infected individuals c_t (cases) is modeled using a discrete renewal process, which is related to the Susceptible–Infected–Recovered model but it is not expressed in differential form. Similar to [7], the time between when a person gets infected and when he subsequently infects other people (i.e. the *serial interval distribution* g) is distributed as

$$g \sim \Gamma(\mu^{\mathcal{G}}, \xi^{\mathcal{G}}).$$

We use as in Flaxman et al. [7] $\mu^{\mathcal{G}} = 6.5$ and $\xi^{\mathcal{G}} = 0.62$ (Fig. 2, upper left), performing sensitivity analysis for other values of $\mu^{\mathcal{G}}$. Note that the distribution of parameter g is fixed in the model.
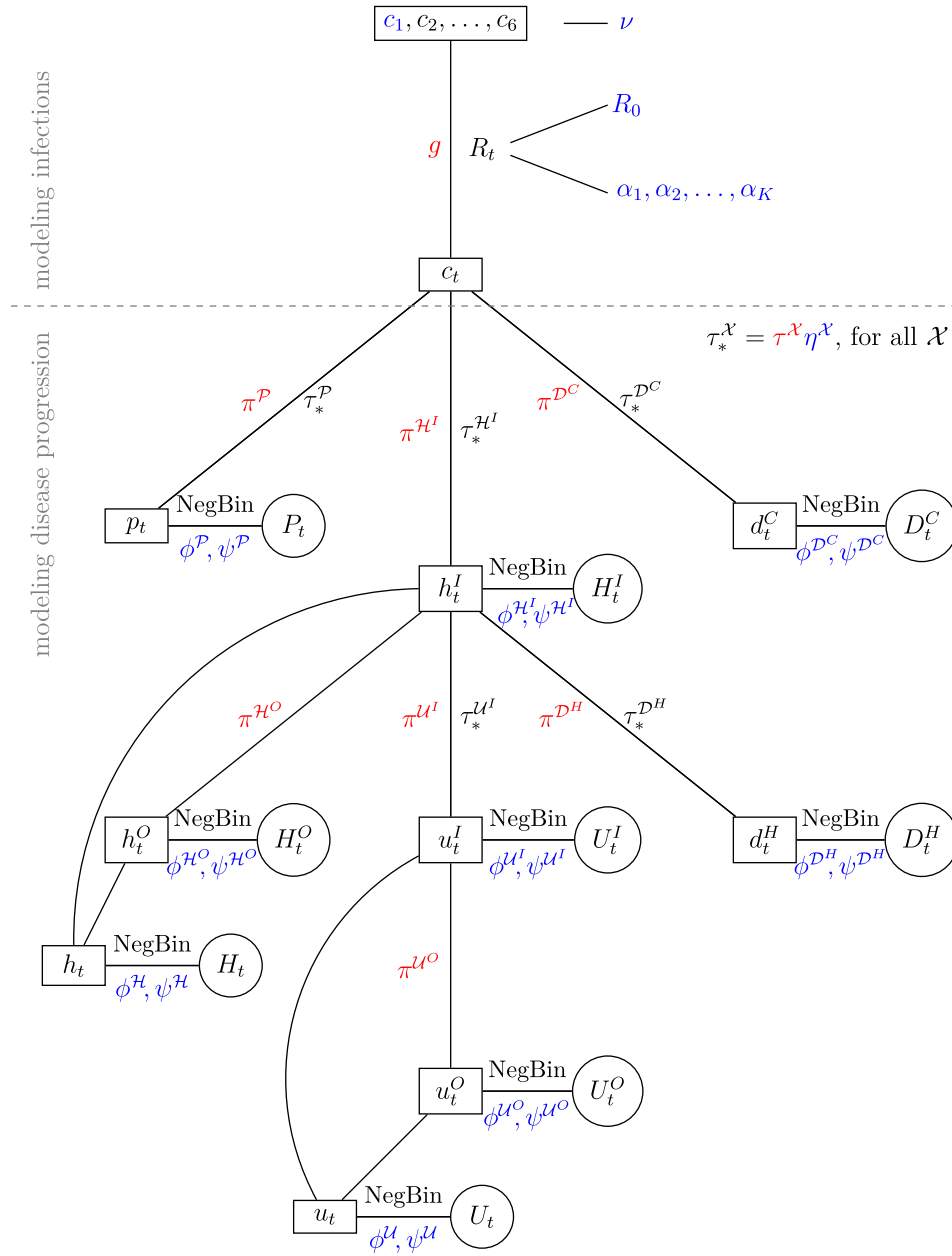
Fig. 1. The model with data values (in circles), its expected values (in boxes), deterministic parameters in red, and 'true' parameters (for which posterior distributions are estimated) in blue (the parameters with prior distributions) or black (the transformed parameters, including expected values in boxes).

The expected number of truly infected individuals c_t on a given day t is then modeled by

$$c_t = \left(1 - \frac{\sum_{k=1}^{t-1} c_k}{N}\right) R_t \sum_{k=0}^{t-1} c_k g_{t-k}, \tag{2}$$

where N is the size of the population, R_t is the time-varying reproduction number modeling the average number of secondary infections over time, and g_s is the discretization of g (1). Infections at time t depend on the number of infections in the previous days, first weighted by the discretized serial interval distribution, and then scaled by R_t and the term accounting for the fact that population can develop (temporal) herd immunity against SARS-CoV-2. The latter would simply be omitted if immunity cannot be developed. Note that we do not have the data for truly infected individuals that would correspond to c_t.

The way the reproduction number varies in time is at the core of our model: we assume that major government changes in policies (for

controlling the COVID-19 epidemic) drive the changes of R_t, where we consider two options. Let $G_1 < G_2 < \cdots < G_K$ be days of major changes of policies. In the first option, where we emulated the study of Flaxman et al. [7], the reproduction number is a piece-wise constant function between days of policy changes:

$$R_t = R_0 \exp\left(-\sum_{k=1}^{K} \alpha_k z_{k,t}\right), \tag{3}$$

where $z_{k,t}$ is equal to 1 if $t \in [G_k, G_{k+1}]$ and 0 otherwise, while $z_{K,t}$ is equal to 1 if $t \geq G_K$ and 0 otherwise. Parameter R_0 is the basic reproduction number with a prior distribution (similarly as in [7])

$$R_0 \sim \mathcal{N}^+(3.28, 0.25). \tag{4}$$

Independent prior distributions for the parameters $\alpha_1, \ldots, \alpha_K$ were chosen to be

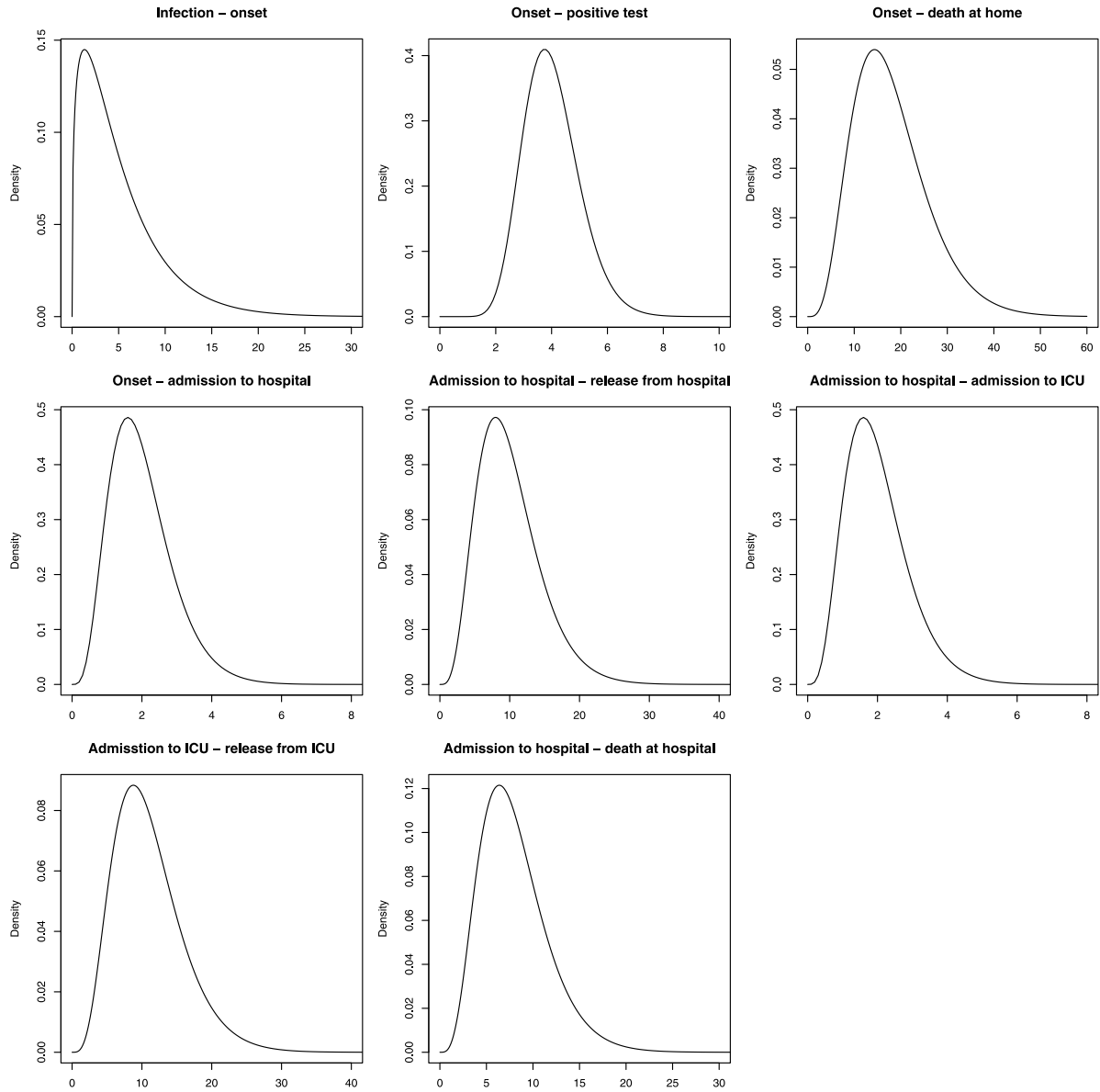$$\alpha_k \sim \text{Exp}(1) - \log(2), \ k = 1, \ldots, K. \tag{5}$$

Fig. 2. Assumed serial interval distribution (g - upper left) and assumed distributions of various times $\pi^{\mathcal{X}}$ used in the model; note different scales.

With these prior distributions (for their graphical presentation see Supplementary Material 1), we have $P(\alpha_k < 0) = 0.5$ and for each $t \geq G_1$

$$\exp\left(-\sum_{k=1}^{K} \alpha_k z_{k,t}\right) \sim \text{Unif}(0, 2). \tag{6}$$

To obtain a continuous estimate of R_t (instead of a piece-wise constant estimate of R_t), we developed a second option: R_t is modeled through

$$R_t = R_0 \exp\left(-\sum_{k=1}^{K+1} \alpha_k s_{k,t}\right), \tag{7}$$

where $s_{k,t}$ is the kth column of the B spline basis matrix for the natural splines, with the positions of the knots set to $G_1 < G_2 < \cdots < G_K$. Natural splines were used in order to maintain low degrees of freedom (note that only one additional parameter needs to be estimated) and to have a continuous function R_t. Priors for R_0 and α_k were chosen as in the first option, but (6) does not hold anymore.

Since the process c_t cannot start from zero, see (2), we have to seed the model. As in [7], we assume that seeding of new infections begins 30 days before observing 10 cumulative deaths. From this date, we seed

the models with six sequential days $c_1, \ldots, c_6$ where we chose

$$c_1 = \cdots = c_6 \sim \text{Exp}(1/\nu) \tag{8}$$

with prior distribution

$$\nu \sim \text{Exp}(0.03). \tag{9}$$

2.1.2. Modeling disease progression

First, we show how confirmed cases P_t (positive cases) are incorporated in the model, others follow a similar logic. We assume that the number of confirmed cases is distributed as

$$P_t \sim \text{NegBin}\left(p_t, p_t + \frac{p_t^2}{\phi^{\mathcal{P}}}\right), \tag{10}$$

where prior distributions were chosen as

$$\phi^{\mathcal{P}} \sim \mathcal{N}^+(0, \psi^{\mathcal{P}}), \quad \psi^{\mathcal{P}} \sim \mathcal{N}^+(8, 2), \tag{11}$$

and p_t is the expected number of confirmed cases that is modeled mechanistically from cases c_t as

$$p_t = \tau_*^{\mathcal{P}} \sum_{k=1}^{t-1} c_k \pi_{t-k}^{\mathcal{P}}. \tag{12}$$

Table 1

Modeling the expected number of individuals at a specific state (first column) from the expected number of individuals at a previous state x_k (third column, see also Fig. 1) via formula (15).

Quantity	Notation	x_k	$\pi^{\mathcal{X}}$	$\tau_*^{\mathcal{X}}$
Positive cases	p_k	c_k	$\Gamma(\mu^{\mathcal{O}}, \xi^{\mathcal{O}}) + \Gamma(\mu^{\mathcal{P}}, \xi^{\mathcal{P}})$	$\tau^{\mathcal{P}} \eta^{\mathcal{P}}$
Hospital in	h_k^I	c_k	$\Gamma(\mu^{\mathcal{O}}, \xi^{\mathcal{O}}) + \Gamma(\mu^{\mathcal{H}^I}, \xi^{\mathcal{H}^I})$	$\tau^{\mathcal{H}^I} \eta^{\mathcal{H}^I}$
Hospital out	h_k^O	h_k^I	$\Gamma(\mu^{\mathcal{H}^O}, \xi^{\mathcal{H}^O})$	1
ICU in	u_k^I	h_k^I	$\Gamma(\mu^{\mathcal{U}^I}, \xi^{\mathcal{U}^I})$	$\tau^{\mathcal{U}^I} \eta^{\mathcal{U}^I}$
ICU out	u_k^O	u_k^I	$\Gamma(\mu^{\mathcal{U}^O}, \xi^{\mathcal{U}^O})$	1
Death in hospitals	d_k^H	h_k^I	$\Gamma(\mu^{\mathcal{D}^H}, \xi^{\mathcal{D}^H})$	$\tau^{\mathcal{D}^H} \eta^{\mathcal{D}^H}$
Death outside the hospitals	d_k^C	c_k	$\Gamma(\mu^{\mathcal{O}}, \xi^{\mathcal{O}}) + \Gamma(\mu^{\mathcal{D}^C}, \xi^{\mathcal{D}^C})$	$\tau^{\mathcal{D}^C} \eta^{\mathcal{D}^C}$

Table 2

Chosen values for deterministic parameters, defined in Table 1, where $\mathcal{O}$ parameters were taken as in Flaxman et al. [7] and others were chosen in exploratory fashion.

$\mathcal{X}$	$\mu^{\mathcal{X}}$	$\xi^{\mathcal{X}}$	$\tau^{\mathcal{X}}$
$\mathcal{O}$	5.1	0.86	
$\mathcal{P}$	4	0.25	0.1
$\mathcal{H}^I$	2	0.45	0.07
$\mathcal{H}^O$	10	0.45	
$\mathcal{U}^I$	2	0.45	0.2
$\mathcal{U}^O$	11	0.45	
$\mathcal{D}^H$	8	0.45	0.2
$\mathcal{D}^C$	18	0.45	0.006

Parameter $\phi^{\mathcal{P}}$ determines the amount of over-dispersion, i.e. the additional variance of the negative binomial distribution above that of the Poisson distribution. Smaller values of the parameter allow for larger over-dispersion. The model could be re-parameterized so that larger values allow for larger over-dispersion, potentially also allowing no over-dispersion (see Supplementary Material 1). Parameter $\tau_*^{\mathcal{P}}$ represents the ratio of cases that are being tested. We model it as

$$\tau_*^{\mathcal{P}} = \tau^{\mathcal{P}} \eta^{\mathcal{P}}, \tag{13}$$

where $\tau^{\mathcal{P}}$ is fixed (chosen in exploratory fashion as a guess of the ratio of cases being tested, see Table 2) and $\eta^{\mathcal{P}}$ is noise around it with prior distribution

$$\eta^{\mathcal{P}} \sim \mathcal{N}^+(1, 0.5). \tag{14}$$

Furthermore, $\pi_s^{\mathcal{P}}$ is discretized distribution for time from infection to a positive test $\pi^{\mathcal{P}}$, which is assumed to be a sum of two independent times: infection to onset (*infection – onset*) and onset to positive test (*onset – positive*)

$$\pi^{\mathcal{P}} \sim \Gamma(\mu^{\mathcal{O}}, \xi^{\mathcal{O}}) + \Gamma(\mu^{\mathcal{P}}, \xi^{\mathcal{P}}).$$

To sum up, the number of confirmed cases p_t is the sum of the past infections c_k weighted by their probability of transition from c_k to p_t. Importantly, $\tau^{\mathcal{P}}$, $\mu^{\mathcal{O}}$, $\xi^{\mathcal{O}}$, $\mu^{\mathcal{P}}$ and $\xi^{\mathcal{P}}$ are fixed, so $\pi^{\mathcal{P}}$ is deterministic. For the infection – onset time we used $\mu^{\mathcal{O}} = 5.1$, $\xi^{\mathcal{O}} = 0.86$ as in [7]. The other deterministic parameters depend on the testing strategies and were chosen in exploratory fashion, the chosen values are given in Table 2.

The data for the number of admissions and releases from hospitals and ICUs, and the number of deaths are incorporated in the model following the same logic as P_t, where we assumed the conditional independence of different data sources given the model parameters (for more details see Section 2.1.3). Hence, the expected values originate from different data sources (see Fig. 1) with different discretized times π and ratios τ_*. Formulas for expected number of individuals at a specific state can be unified through

$$\tau_*^{\mathcal{X}} \sum_{k=1}^{t-1} x_k \pi_{t-k}^{\mathcal{X}}, \tag{15}$$

where x_k is an appropriate expected number of individuals at a previous state (see Fig. 1), and the meaning of $\tau_*^{\mathcal{X}}$ and $\pi^{\mathcal{X}}$ remains the same. To be succinct, we present the rest of the model in Table 1. Note that column $\tau_*^{\mathcal{X}}$ is fixed to 1 for hospital out and ICU out. This holds when H_t^O is just time-lagged H_t^I, meaning that this data source does not separate deceased patients from cured. The chosen values for the deterministic parameters are given in Table 2. The assumed distributions of various times used in our analyses are shown in Fig. 2.

Expected number of hospitalized patients h_t and number of patients in ICUs are modeled differently compared to others. We used

$$\begin{aligned} h_1 &= h_1^I - h_1^O, \\ h_t &= h_{t-1} + h_t^I - h_t^O, \quad t = 2, \ldots, n, \end{aligned} \tag{16}$$

and similarly for u_t. We use data on H_t in addition to H_t^I and H_t^O due to some inconsistencies in our data sources between the three and we did

similar for ICU. Note also that we do not have a possibility to go from u_t^I to 'death from ICUs' as there is no such data available for Slovenia.

Daily number of deaths d_t is then forecasted using $d_t = d_t^C + d_t^H$. Since the daily number of deaths D_t was consistent in our data, i.e. $D_t = D_t^C + D_t^H$, additional modeling of D_t was not necessary.

2.1.3. Model specification

According to Bayes theorem, the joint posterior distribution $p(\theta|\mathcal{D})$ is proportional to $p(\mathcal{D}|\theta)p(\theta)$ where $p(\mathcal{D}|\theta)$ is the likelihood of the data $\mathcal{D}$ and $p(\theta)$ is the joint prior distribution for all the model parameters θ. Specifically, the vector of parameters is

$$\begin{aligned} \theta = (\{p_t, h_t, h_t^I, h_t^O, u_t, u_t^I, u_t^O, d_t^H, d_t^C\}_{t=1}^n, \{\phi^{\mathcal{X}}, \psi^{\mathcal{X}}, \tau_*^{\mathcal{X}}, \eta^{\mathcal{X}}\}_{\mathcal{X}}, \\ \{c_t\}_{t=1}^n, \{R_t\}_{t=1}^n, \{\alpha_k\}_{k=1}^K, R_0, \nu), \end{aligned}$$

where $\mathcal{X}$ goes through all data sources, i.e. $\mathcal{P}$ (positive cases), $\mathcal{H}$ (hospital), $\mathcal{H}^I$ (hospital in), $\mathcal{H}^O$ (hospital out), $\mathcal{U}$ (ICU), $\mathcal{U}^I$ (ICU in), $\mathcal{U}^O$ (ICU out), $\mathcal{D}^H$ (death in hospitals) and $\mathcal{D}^C$ (death outside hospitals). The data are

$$\mathcal{D} = \left(\{P_t, H_t, H_t^I, H_t^O, U_t, U_t^I, U_t^O, D_t^H, D_t^C\}_{t=1}^n \right).$$

We assume conditional independence of different data sources given the model parameters, so the likelihood can be expressed as

$$p(\mathcal{D}|\theta) = \prod_{t=1}^n p(P_t|p_t, \phi^{\mathcal{P}}) \cdot \ldots \cdot \prod_{t=1}^n p(D_t^C|d_t^C, \phi^{\mathcal{D}^C}),$$

where $p(P_t|p_t, \phi^{\mathcal{P}})$ is the probability density function of $\text{NegBin}(p_t, p_t + p_t^2/\phi^{\mathcal{P}})$ as specified in (10), similarly for others.

We assume independent prior distributions if not stated otherwise in the previous subsections. In particular, the joint prior distribution is

$$p(\theta) = \left(\prod_{\mathcal{X}} p(\phi^{\mathcal{X}}|\psi^{\mathcal{X}}) p(\psi^{\mathcal{X}}) p(\eta^{\mathcal{X}}) \right) \left(\prod_{k=1}^K p(\alpha_k) \right) p(R_0) p(c_1|\nu) p(\nu),$$

where the distributions of priors, from left to right, are specified in (11), (14), (5), (4), (8) and (9) whereas all other parameters of θ that are not specified in the above joint prior are transformed parameters via (2), (3) or (7), (12), (13), (15) and (16).

In addition to the posterior distributions for all the parameters, including the expected numbers of individuals at all states (x_t), the outputs of our model are also their forecasts for a specified number of days based on the posterior predictive distribution (the reliability of forecasts is validated using methods presented in Section 2.4).

2.2. Possible extensions and modifications of the model

The described model could be extended to take into account data from different countries $m \in \{1, 2, \ldots, M\}$ (or different regions etc.) and make estimation and forecasts at country level. A possible generalization is presented in Fig. 3. Note that there are three options for modeling each parameter, for brevity we only describe them for $\phi^{\mathcal{P}}$. The first option is to assume that all countries have the same parameter $\phi^{\mathcal{P}}$, which could lead to a poor fit due to using only one parameter. The second option is that each country has its own $\phi_m^{\mathcal{P}}$ and that they
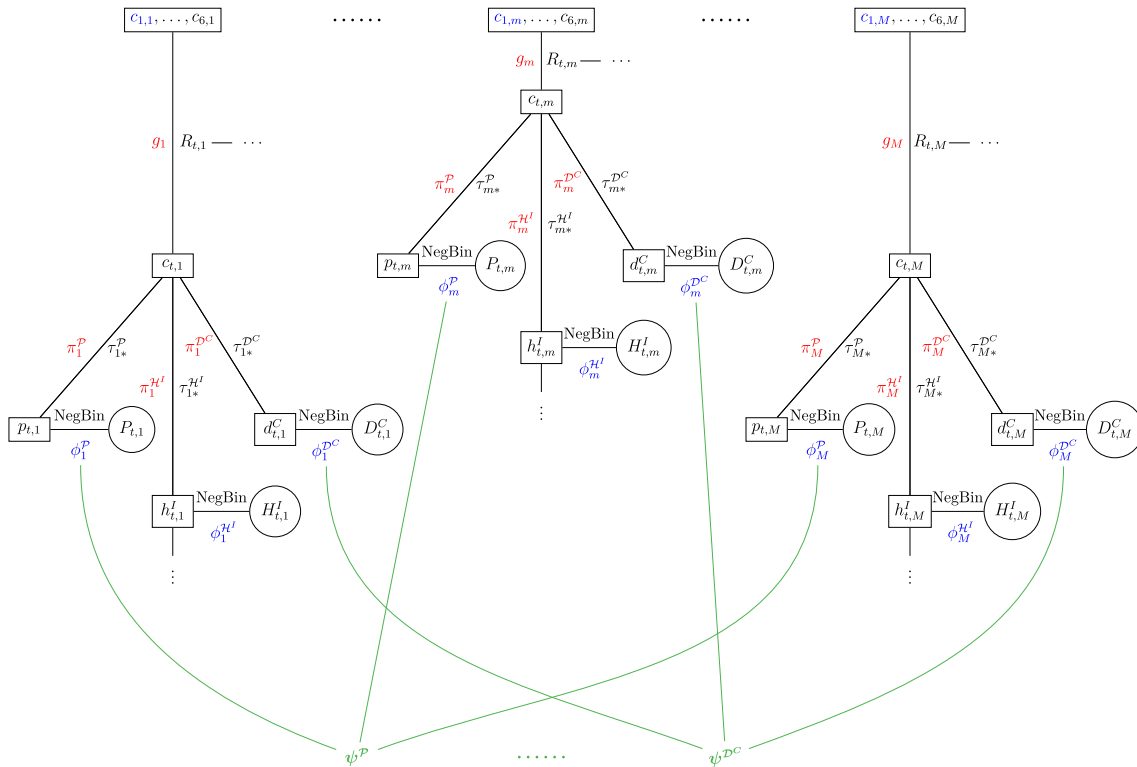
Fig. 3. A possible extension of the core model (Fig. 1), where hyperparameters are colored green.

are independent of each other, i.e. giving each $\phi_m^{\mathcal{P}}$ its own prior, which could lead to overfitting problems due to many parameters. The third option is to model this parameter hierarchically as in Fig. 3, i.e.

$$\phi_1^{\mathcal{P}}, \phi_2^{\mathcal{P}}, \ldots, \phi_M^{\mathcal{P}} \sim \mathcal{N}^+(0, \psi^{\mathcal{P}})$$

with a prior distribution for $\psi^{\mathcal{P}}$. In this way, we allow for the heterogeneity of parameters across countries, while using a common distribution with hyperparameter $\psi^{\mathcal{P}}$ to introduce some dependence and constrain them in order to avoid the overfitting problem.

Furthermore, the model could be easily modified to correspond to the availability and reliability of the data collected in a specific country as certain branches of the model presented in Figs. 1 and 3 could be removed. In fact, this is why the expected number of patients admitted to hospitals h_t^I and expected number of deaths outside of hospitals d_t^C are modeled directly from the expected number of truly infected c_t, instead through positive cases p_t. Namely, data on positive cases for the COVID-19 pandemic are often unreliable and depend strongly upon testing strategy which changes during the course of the pandemic and substantially varies between different countries. In this way, the forecasts of our model are not driven by the number of positive cases, which, if deemed necessary, can also simply be excluded from the model.

2.3. Model stratified by hospitals

Here we show a modification of our model to incorporate data on hospital level. The model is presented in Fig. 4.

Infections are modeled as before, then separate submodels for hospitals are connected to c_t. Data for M hospitals are denoted by $H_{t,m}$, $H_{t,m}^I$, $H_{t,m}^O$, $m = 1, \ldots, M$, similarly for ICU data. For each hospital we use the same model as in Section 2.1.2, where the parameters $\tau_{m*}^{\mathcal{H}^I}$ (representing the probability of admission to hospital m) are hospital specific, whereas all other parameters are not hospital specific, see Fig. 4. We opted for this assumption in our application to Slovene data for two reasons: (1) the daily numbers for hospitals were small

(increasing the variability of our estimates), and (2) the organization of public health care in Slovenia is unified across the entire country, COVID-19 patients are treated only in public hospitals. This assumption could be easily relaxed in the model, if needed.

For Slovene hospitals we chose $\tau_m^{\mathcal{H}^I} = 0.07 \cdot w_m$ (in the core model $\tau^{\mathcal{H}^I} = 0.07$ was used, see Table 2) where the weights w_m were determined from the data, yielding $w_1 = 0.43$, $w_2 = 0.28$, $w_3 = 0.13$ and $w_4 = 0.16$ for University Medical Centre (UMC) Ljubljana, UMC Maribor, University Clinic (UC) Golnik and General Hospital (GH) Celje, respectively (COVID-19 patients in Slovenia have been admitted only to these four hospitals).

2.4. Computational aspects and model validation

Convergence of the models was assessed based on $\hat{R}$ statistics [23], trace plots, estimated posteriors, mean with 50%, and 90% credible intervals for R_0, R_n (the last day of studied period) and the α parameters for each chain individually. Effective sample size is also reported.

Sensitivity analysis was performed for mean parameter of serial distribution $\mu^{\mathcal{G}}$ considering 6, 6.5, 7 and 8 (see Supplementary Material 2 and 3). Folded-normal distribution was compared with other commonly used prior distributions for positive quantities, i.e., gamma and log-normal distributions (see Supplementary Material 1). Several choices for the over-dispersion parameter of the assumed negative binomial distributions were considered (see Supplementary Material 1). The differences were not substantial.

The model presented in Section 2.1 is also compared with the Flaxman model [7] using Slovene data, showing that the estimate of R_t for Slovenia after the final intervention based on our model is larger with slightly narrower CI (see Supplementary Material 1 for more details). The effect of including several data sources in our model is shown in Supplementary Material 1; using only a single data source, i.e. data on the daily number of deaths results in wider CIs.

For each model we evaluate the quality of out-of-time forecasts for F future days as follows. It was decided that forecasts will not be
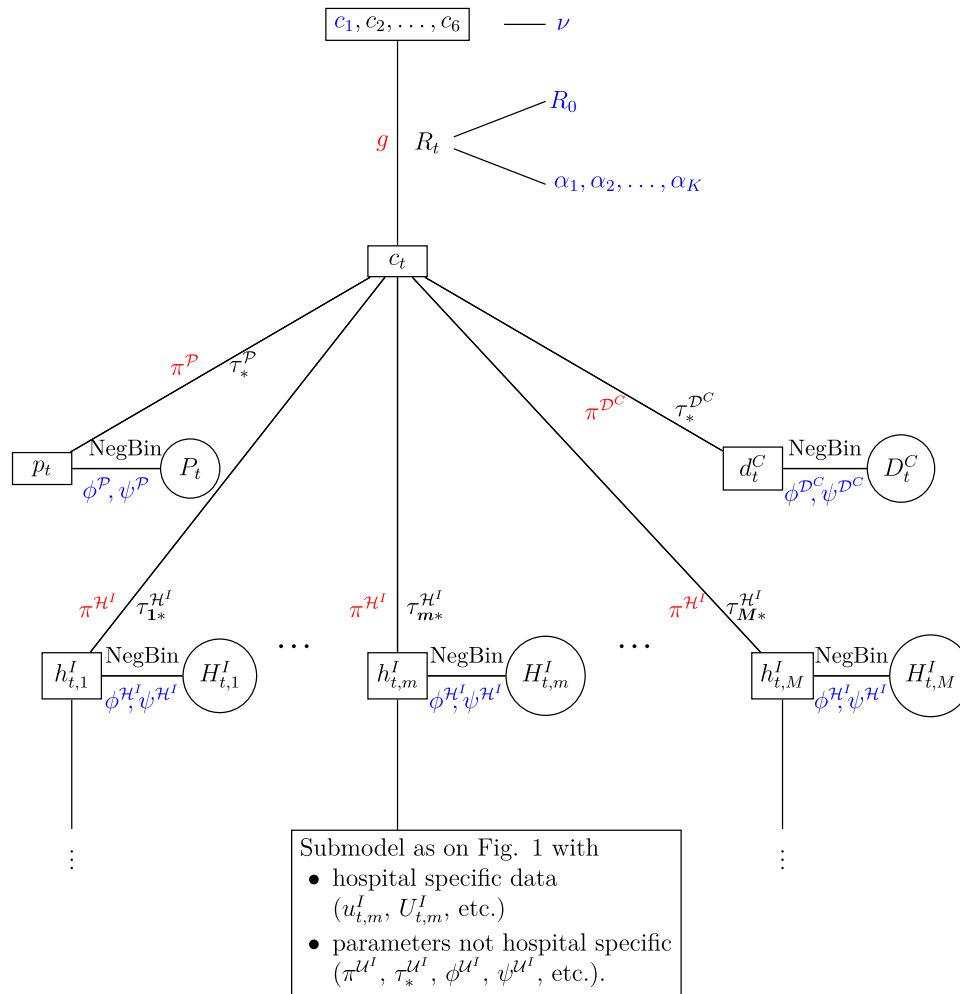
Fig. 4. Model stratified by hospitals.

made until the occurrence of 10 cumulative deaths, which occurred in Slovenia on March 29, 2020; the models were then refitted daily up to and including May 24, 2020, evaluating the quality of the out-of-time forecasts for $F = 1, 2, \ldots, 10$. The refitted models were used to compute the exact expected log predictive density (EELPDS) [24] separately for each variable (deaths, hospital, ICU, etc.) and F with larger values of EELPDS implying better predictive ability.

The models were fitted in R [25] (R version 3.6.3) using R package **rstan** [26]. The models from May 6, 2020 onwards are fitted using 1600 iterations (800 warmup iterations), 4 chains, keeping every 4th sample, setting the target average proposal acceptance probability during standard adaptation period to 0.98 and putting a cap on the depth of the trees that it evaluates during each iteration to 15. The earlier models were fitted using 2600 iterations (1300 warmup iterations) setting the target average proposal acceptance probability to 0.995 with the other settings as described before. All the code needed to reproduce the results is available on GitHub [27]. The analysis was performed on cluster of CentOS based containers.

3. Application to Slovene COVID-19 data

In this section, we apply the proposed modeling framework to publicly available Slovene COVID-19 data [22]. We have analyzed the data up to and including June 3, 2020. For a brief presentation of the data see Fig. 5 where for the study period we plot the number of patients treated in hospitals and ICUs on a given date, cumulative number of deaths and cumulative number of deaths occurring in hospitals. First

infection was confirmed on March 4, 2020 and in total there were 1477 infections in the studied period; the largest number of confirmed infections in a single day occurred on March 26 when 61 new infections were confirmed. First death due to COVID-19 occurred on March 14, 2020; in total there were 109 deaths, out of these 55 (50.5%) occurring in the hospitals. First patients were admitted to hospitals and ICUs on March 11, 2020; in total 352 and 73 patients were admitted to hospitals and ICUs, respectively, with the largest number of new admissions occurring on March 27, 2020 and March 25, 2020, when 23 and 6 new patients were admitted to hospitals and ICUs, respectively.

In order to control the epidemic, Slovene Government banned public events on March 10, 2020, which was followed by implementation of complete lock-down on March 20, 2020 and prohibition of movement outside of the municipality of residence on March 30, 2020. The government started to relax the restrictions on April 30, 2020. Correspondingly, $G_1 < G_2 < G_3 < G_4$ were set to these respective dates when modeling R_t.

3.1. Results for the model at the national level

The first goal was to estimate R_t and the number of infected individuals, and make forecasts for, among others, the number of hospitalized patients and number of patients in ICUs using data at country level. For this purpose, we use the model from Section 2.1. A series of selected results is presented in Fig. 6; complete results are available as Supplementary Material 2.

Estimated number of infected individuals and basic reproduction number R_0 were very similar when modeling R_t using splines and
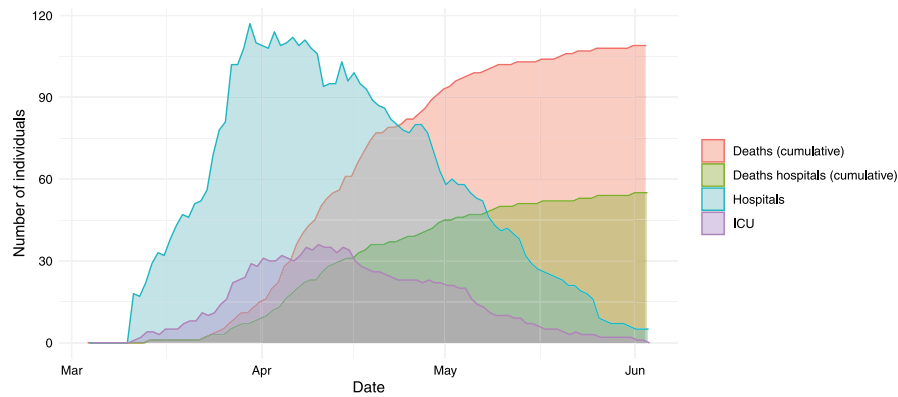
Fig. 5. Number of patients treated in hospitals and ICUs on a given date, cumulative number of deaths and cumulative number of deaths occurring in hospitals in Slovenia, March 4–June 3, 2020.
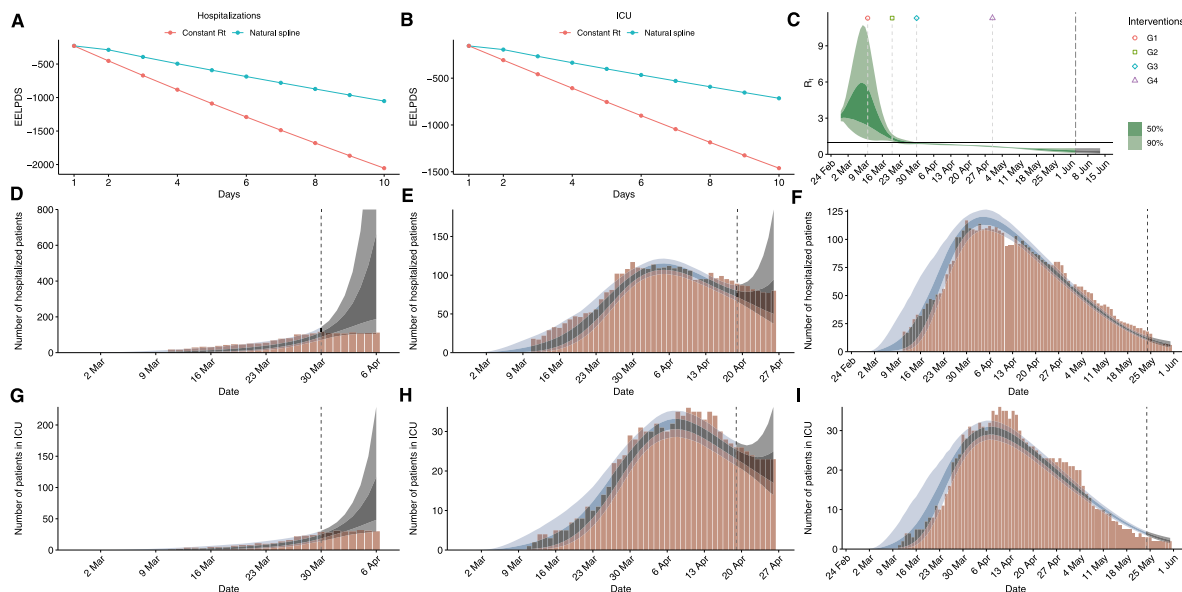

Fig. 6. First row: EELDPS values for number of hospitalized patients (A) and in ICU (B) based on models that are fitted with data until May 24, 2020; estimated 50% and 90% CI for the R_t of SARS-CoV-2 for Slovenia using natural spline with marked days of NPIs (C). Second row shows 7-day forecasts for hospitalized patients and third row for ICU patients using natural spline. The observed number of patients (brown) are shown together with the respective predicted values and 7-day forecasts (darker and lighter shaded blue areas represent 50% and 90% CI, respectively); results are shown for fitted models until March 30 (D and G), April 19 (E and H), May 24 (F and I).

piece-wise constant function (see Supplementary Material 2). We have estimated that 7287 people (0.350%; 90% credible interval (CI): [0.245 –0.573]%) have been infected with SARS-CoV-2 throughout the study period when modeling with splines. The basic reproduction number has increased from 3.17 (90% CI [2.74–3.59]) to 3.92 (90% CI [1.56–9.68]) until the adoption of measures to control the epidemic, after which the effective R_t started to decrease reaching its lowest value of 0.17 (90% CI [0.05–0.51]) at the end of the study period (see also Fig. 6, panel C). Based on these results it can be concluded that the adopted NPIs in Slovenia were effective in slowing down the spread of the epidemic which eventually resulted in the end of the first wave. Using piece-wise constant function in the model, the largest and smallest estimated R_t were less extreme, 3.21 [1.00–5.81] and 0.48 [0.44–0.52], respectively. This is due to a supposition that R_t does not change between interventions and can therefore be only interpreted as *average* R_t in the period between two successive interventions, whereas when using splines R_t can be interpreted on a daily basis.

Based on the estimates of the number of infected individuals and estimated number of deaths, we estimate, using splines when modeling R_t, that throughout the study period infection fatality rate (IFR) was

1.56% (90% CI [0.94–2.21]%). In Slovenia, the share of estimated deaths outside the hospitals for this period was large, 47.15% (90% CI [38.5–55.26]%), due to deaths in the retirement homes, since there were many outbreaks therein. Excluding deaths not occurring in the hospitals results in the IFR of 0.8% (90% CI [0.48–1.26]%). It is also estimated that throughout the study period the proportion of unidentified cases (defined as 1 minus the ratio of the number of positive and infected cases), which can account for asymptomatic cases, was equal to 88% (90% CI [83-93]%).

Fig. 6 shows 3 examples of models fitted to various end-dates. The quality of the out-of-time forecasts as measured by EELDPS was better when using splines, with good ability to make out-of-time forecasts throughout the course of the Slovene COVID-19 epidemic (see Fig. 6, panels A, B). Considering only early data, the longer the interval between end-date and final time-point prediction the wider the CI, often even being uninformative (see Fig. 6, panels D, G). As expected, when more data are available the precision of the forecasts improves with much narrower CIs (see Fig. 6, panels E, F, H, I).

It is also important to note that the estimates of R_t at the final time-point can be imprecise when the final time-point is near the last
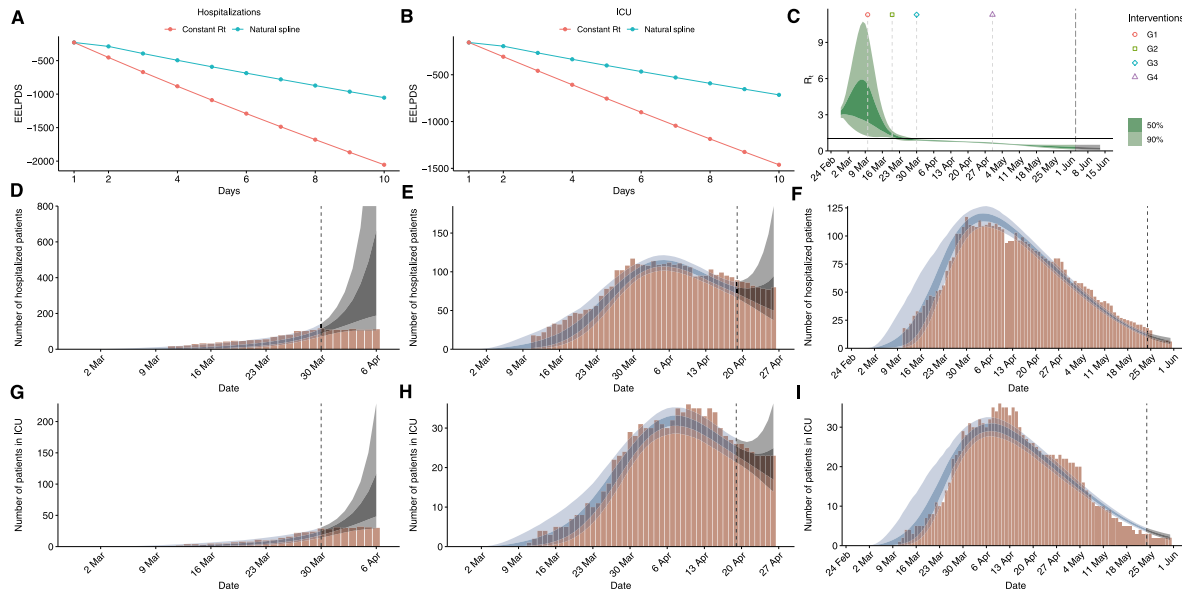
Fig. 7. First row: EELPDS values for number of hospitalized patients in UMC Ljubljana (A) and UMC Maribor (B) based on models that are fitted with data until May 24, 2020; estimated 50% and 90% CI for the R_t of SARS-CoV-2 for Slovenia using natural spline with marked days of NPIs (C). Second row shows 7-day forecasts for hospitalized patients in UMC Ljubljana and third row in UMC Maribor using natural spline. The observed number of patients (brown) are shown together with the respective predicted values and 7-day forecasts (darker and lighter shaded blue areas represent 50% and 90% CI, respectively); results are shown for fitted models until March 30 (D and G), April 19 (E and H), May 24 (F and I).

intervention. This is a natural consequence of having few data when estimating α_K. Therefore, a long enough interval between the interventions and after the latest intervention is required in order to obtain informative estimates. For example, final R_t estimates as given by 90% CI when using data up to and including May 1 (one day after some restrictions were lifted), May 5 and May 10 were [0.11–2.47], [0.05–1.40] and [0.03–0.61], respectively, when using splines and similarly (although less precise with wider CIs) when using piece-wise constant function (see Supplementary Material 2).

3.2. Results for model at the hospital level

Here, we use the data on hospital level with the goal of estimating R_t and the number of infected individuals and to forecast the number of COVID-19 patients in each hospital. We use the model presented in Section 2.3. A series of selected results are presented in Fig. 7; complete results appear as Supplementary Material 3.

As with the analysis at the national level, using splines to model the reproduction number resulted in larger EELPDS providing reasonable short-term forecasts (for time intervals $\leq$ 7 days) for each hospital throughout the course of the Slovene COVID-19 epidemic (see Fig. 7, panels A, B). A similar time trend is estimated for the reproduction number as in the analysis at national level (see Fig. 7, panel C), providing similar estimates of the total number of infected individuals. As expected, since essentially no new information was introduced into the model, the precision of the estimates is very similar as in the previous analysis.

4. Discussion

The idea behind the model of Flaxman was to improve the accuracy of estimates (i.e., to narrow its credible intervals) with pooling information across different countries while only using data on number of deaths. To achieve this, several strong assumptions had to be made. For instance, they had to assume that the effect of each NPI was the same across countries. Our model uses several data sources from the same country to make estimates as accurate as possible without making such strong assumptions. Using Slovene data, we have illustrated the

gain in efficiency by using several data sources in comparison with using a single data source, i.e. daily number of deaths. The modeling framework can be adapted to use data at different levels of aggregation, as was demonstrated on the Slovene data when making forecasts at the hospital level. As expected, in our application this did not further improve the efficiency of our estimates since essentially no new information was added in the model. However, having separate forecasts for each hospitals was important from a practical perspective for monitoring the state of the Slovene health-care system.

With semiparametric model, there are several parameters which need to be specified and are influencing the estimates. Some of these parameters are obtained from the literature, the others can be estimated from patient level data, if available, and plugged into the model, e.g., the duration of time spent in hospital and ICU. However, if some parameters cannot be determined, the model can be simplified by omitting these parameters along with the data connected to them.

The model estimates and forecasts are largely driven by data, hence their accuracy relies heavily on the quality of data. If some data sources are not reliable enough or are not available then the model can be simplified by omitting them. We have explained how sharing and/or pooling information across different countries and/or regions could be done, similarly as in Flaxman et al. [7], but this extension has to be used with caution since it adds new assumptions to the model, which are difficult to verify. The extended model would have to be thoroughly validated.

When applying the proposed framework to Slovene data, we have estimated that R_t reduced after adopting the NPIs showing their efficacy in controlling the transmission of SARS-CoV-2. However, after adopting all NPIs, R_t as estimated by our model was larger than when using the Flaxman model [7], suggesting that for the case of Slovene data, using the data from other countries would result in overestimation of the effect of the NPIs on the reduction of R_t. This is also in line with our previous results [21], where we have observed that excluding the two of the most affected European countries (Italy and Spain) from the Flaxman model, results in much smaller estimated reduction of R_t after adopting the NPIs in Slovenia. The effects of the NPIs seem to be the largest in these two countries.

Based on the proposed model we have estimated the proportion of cumulatively infected people (attack rate) of 0.350% (90% credible

interval (CI): [0.245–0.573]%) of the Slovene population suggesting that Slovenia had one of the smallest attack rates in Europe. At the same time, we have estimated the infection fatality rate (IFR) of 1.56% (90% CI [0.94–2.21]%). While estimating the IFR is difficult, especially for COVID-19, other studies tend to report it around 0.5%–1% [28], which implies a large estimated IFR for Slovenia. This can be explained by noting that during the first wave of Slovene epidemic, the virus transmitted largely among older people inside retirement homes who are at much higher risk of dying due to COVID-19. Up to May 4, around 80% of all deaths occurred among people from retirement homes [29] and all deaths outside the hospitals were in retirement homes. Excluding the estimated number of deaths occurring outside the hospitals yields a much smaller IFR of 0.80% (90% CI [0.48–1.26]%) thus more closely agreeing with other studies. It is of importance therefore to prevent the spread of the virus among elderly.

The estimated percentage of unidentified cases based on Slovene data was equal to 88% (90% CI [83-93]%). Unidentified cases can to some extent be attributed to asymptomatic or mild symptomatic cases but could also reflect testing strategy of a country. Asymptomatic cases can transmit the virus [30] causing difficulties in the control of the epidemic [31]. Other studies have estimated that 40% to 45% of those infected with SARS-CoV-2 remain asymptomatic [32] with large differences between the studies (e.g., 17.9% in [33] to 87.9% in [34]). A large estimate for Slovenia might be a consequence of the fact that people with mild symptoms during the first wave were instructed to self-isolate and were in large majority not tested for SARS-CoV-2.

5. Conclusion

We have presented a semiparametric framework for modeling and forecasting COVID-19 pandemic combining Bayesian inference with current epidemiological knowledge about SARS-CoV-2. It is based on the model set by Flaxman et al. [7], whilst giving important extensions and modifications in the model. The presented framework can be used for estimating the number of individuals infected with SARS-Cov-2, its effective reproduction number, infection fatality rate, proportion of cumulatively infected people and proportion of unidentified cases, as well as for accurate short term forecasts of various important indicators of the state of epidemic for the healthcare system, e.g. the daily number of patients admitted to hospitals and ICUs. Furthermore, we have shown how the proposed framework could be extended to more countries/regions, thus allowing for comparison between them.

Applying the framework to the Slovene data we found, that the adoption of NPIs was effective in limiting the first wave in Slovenia which mirrored in small estimated percentage of cumulatively infected people of 0.350%, 90% CI [0.245–0.573]%. Despite that finding, the estimated infection fatality rate for Slovenia was high (1.56%, 90% CI [0.94–2.21]%) which can be explained by entries of the virus into retirement homes. When eliminating some of the deaths that could be spared if there were no infections in retirement homes, an estimate for IFR would fall to 0.80% (90% CI [0.48–1.26]%).

The proposed models can be easily implemented by those familiar with Bayesian inference. The R code of our model applied to Slovene data, which is partially based on the freely available R code provided by Flaxman et al. [7], is available on GitHub [27].

CRediT authorship contribution statement

Damjan Manevski: Formal analysis, Software, Writing - review & editing. **Nina Ružić Gorenjec:** Methodology, Formal analysis, Writing - review & editing. **Nataša Kejžar:** Formal analysis, Writing - review & editing. **Rok Blagus:** Conceptualization, Methodology, Formal analysis, Software, Writing - original draft, Supervision, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix A. Supplementary materials

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.mbs.2020.108466.

References

[1] K. Rothman, S. Greenland, T. Lash, Modern Epidemiology, Wolters Kluwer Health/Lippincott Williams & Wilkins, 2015.

[2] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, J. Shaman, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2), Science 368 (6490) (2020) 489–493, http://dx.doi.org/10.1126/science.abb3221.

[3] Z. Du, L. Wang, S. Cauchemez, X. Xu, X. Wang, B.J. Cowling, L.A. Meyers, Risk for transportation of coronavirus disease from Wuhan to other cities in China, Emerg. Infect. Dis. 26 (5) (2020) 1049, http://dx.doi.org/10.3201/eid2605.200146.

[4] J.T. Wu, K. Leung, G.M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, Lancet 395 (10225) (2020) 689–697, http://dx.doi.org/10.1016/S0140-6736(20)30260-9.

[5] N. Imai, I. Dorigatti, A. Cori, S. Riley, N. Ferguson, Estimating the Potential Total Number of Novel Coronavirus (2019-nCoV) Cases in Wuhan City, China, Preprint published by the Imperial College London, 2020.

[6] J. Zhang, M. Litvinova, Y. Liang, Y. Wang, W. Wang, S. Zhao, Q. Wu, S. Merler, C. Viboud, A. Vespignani, et al., Age profile of susceptibility, mixing, and social distancing shape the dynamics of the novel coronavirus disease 2019 outbreak in China, medRxiv (2020) http://dx.doi.org/10.1101/2020.03.19.20039107.

[7] S. Flaxman, S. Mishra, A. Gandy, H. Unwin, T. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J. Eaton, M. Monod, A. Ghani, C. Donnelly, S. Riley, M. Vollmer, N. Ferguson, L. Okell, S. Bhatt, P. Perez-Guzman, P. Walker, Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe, Nature 584 (2020) 257–261, http://dx.doi.org/10.1038/s41586-020-2405-7.

[8] J. Riou, C.L. Althaus, Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020, Eurosurveillance 25 (4) (2020) 2000058, http://dx.doi.org/10.2807/1560-7917.ES.2020.25.4.2000058.

[9] D. Fanelli, F. Piazza, Analysis and forecast of COVID-19 spreading in China, Italy and France, Chaos Solitons Fractals 134 (2020) 109761, http://dx.doi.org/10.1016/j.chaos.2020.109761.

[10] L. Zhong, L. Mu, J. Li, J. Wang, Z. Yin, D. Liu, Early prediction of the 2019 novel coronavirus outbreak in the mainland China based on simple mathematical model, IEEE Access 8 (2020) 51761–51769, http://dx.doi.org/10.1109/ACCESS.2020.2979599.

[11] T.C. Germann, K. Kadau, I.M. Longini, C.A. Macken, Mitigation strategies for pandemic influenza in the United States, Proc. Natl. Acad. Sci. 103 (15) (2006) 5935–5940, http://dx.doi.org/10.1073/pnas.0601266103, arXiv:https://www.pnas.org/content/103/15/5935.full.pdf.

[12] Q. Li, B. Tang, N.L. Bragazzi, Y. Xiao, J. Wu, Modeling the impact of mass influenza vaccination and public health interventions on COVID-19 epidemics with limited detection capability, Math. Biosci. 325 (2020) 108378, http://dx.doi.org/10.1016/j.mbs.2020.108378.

[13] M.A. Acuña Zegarra, M. Santana-Cibrian, J.X. Velasco-Hernandez, Modeling behavioral change and COVID-19 containment in Mexico: A trade-off between lockdown and compliance, Math. Biosci. 325 (2020) 108370, http://dx.doi.org/10.1016/j.mbs.2020.108370.

[14] H. Zhao, Z. Feng, Staggered release policies for COVID-19 control: Costs and benefits of relaxing restrictions by age and risk, Math. Biosci. 326 (2020) 108405, http://dx.doi.org/10.1016/j.mbs.2020.108405.

[15] Ž. Zaplotnik, A. Gavrić, L. Medic, Simulation of the COVID-19 epidemic on the social network of Slovenia: estimating the intrinsic forecast uncertainty, PLOS ONE 15 (8) (2020) 1–22, http://dx.doi.org/10.1371/journal.pone.0238090.

[16] D. Balcan, B. Gonçalves, H. Hu, J.J. Ramasco, V. Colizza, A. Vespignani, Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model, J. Comput. Sci. 1 (3) (2010) 132–145, http://dx.doi.org/10.1016/j.jocs.2010.07.002.

[17] T. Kypraios, P. Neal, D. Prangle, A tutorial introduction to Bayesian inference for stochastic epidemic models using Approximate Bayesian Computation, Math. Biosci. 287 (2017) 42–53, http://dx.doi.org/10.1016/j.mbs.2016.07.001.

[18] P.D. O'Neill, A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods, Math. Biosci. 180 (1) (2002) 103–114, http://dx.doi.org/10.1016/S0025-5564(02)00109-8.

[19] C. Hsu, A. Yen, L. Chen, H. Chen, Analysis of household data on influenza epidemic with Bayesian hierarchical model, Math. Biosci. 261 (2015) 13–26, http://dx.doi.org/10.1016/j.mbs.2014.11.006.

[20] N. Becker, T. Britton, Statistical studies of infectious disease incidence, J. R. Stat. Soc. Ser. B Stat. Methodol. 61 (2) (1999) 287–307, http://dx.doi.org/10.1111/1467-9868.00177.

[21] D. Manevski, M. Pohar Perme, R. Blagus, Estimation of the reproductive number and the outbreak size of SARS-CoV-2 in Slovenia, Slov. Med. J. 89 (2020) 1–12, http://dx.doi.org/10.6016/ZdravVestn.3068.

[22] COVID-19 sledilnik Team, COVID-19 sledilnik, 2020, URL https://covid-19.sledilnik.org, (Accessed 5 June 2020).

[23] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, D. Rubin, Bayesian Data Analysis, Chapman & Hall/CRC Texts in Statistical Science, CRC Press, 2013.

[24] A. Vehtari, A. Gelman, J. Gabry, Practical bayesian model evaluation using leave-one-out cross-validation and WAIC, Stat. Comput. 27 (2017) 1413–1432, http://dx.doi.org/10.1007/s11222-016-9696-4.

[25] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2020, URL https://www.R-project.org/.

[26] Stan Development Team, RStan: The R Interface To Stan, R Package Version 2.19.3, 2020, URL http://mc-stan.org/.

[27] Modelling COVID-19 pandemic using Bayesian analysis with application to Slovenian data, URL https://github.com/damjanm/COVID-19-Slovenia.

[28] S. Mallapaty, How deadly is the coronavirus? Scientists are close to an answer, Nature 582 (2020) 467–468, http://dx.doi.org/10.1038/d41586-020-01738-2.

[29] T. Pihlar, Skoraj osemdeset odstotkov vseh smrti, povezanih s koronavirusno boleznijo, je bilo v domovih starejših, Dnevnik (2020) URL https://www.dnevnik.si/1042928747.

[30] Y. Bai, L. Yao, T. Wei, F. Tian, D.-Y. Jin, L. Chen, M. Wang, Presumed asymptomatic carrier transmission of COVID-19, JAMA 323 (14) (2020) 1406–1407, http://dx.doi.org/10.1001/jama.2020.2565.

[31] Q.-X. Long, X.-J. Tang, Q.-L. Shi, Q. Li, H.-J. Deng, J. Yuan, J.-L. Hu, W. Xu, Y. Zhang, F.-J. Lv, K. Su, F. Zhang, J. Gong, B. Wu, X.-M. Liu, J.-J. Li, J.-F. Qiu, J. Chen, A.-L. Huang, Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections, Nat. Med. 26 (2020) 1200–1204, http://dx.doi.org/10.1038/s41591-020-0965-6.

[32] D.P. Oran, E.J. Topol, Prevalence of asymptomatic SARS-CoV-2 infection, Ann. Internal Med. (2020) http://dx.doi.org/10.7326/M20-3012.

[33] K. Mizumoto, K. Kagaya, A. Zarebski, G. Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the diamond princess cruise ship, Yokohama, Japan, 2020, Eurosurveillance 25 (10) (2020) http://dx.doi.org/10.2807/1560-7917.ES.2020.25.10.2000180.

[34] T.P. Baggett, H. Keyes, N. Sporn, J.M. Gaeta, COVID-19 outbreak at a large homeless shelter in boston: Implications for universal testing, medRxiv (2020) http://dx.doi.org/10.1101/2020.04.12.20059618, arXiv:https://www.medrxiv.org/content/early/2020/04/15/2020.04.12.20059618.full.pdf.